On the Diameter of Sensor Networks

Esther H. Jennings and Clayton M. Okino
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
{jennings,cokino}@arcadia.jpl.nasa.gov

Abstract—To prolong the life-time of sensor nodes used in space exploration, cooperative modulation techniques had been proposed which rely on information sharing among the nodes, e.g. broadcasting. The desire to efficiently reduce the overall energy-per-bit of a node motivated this study since the diameter directly affects broadcasting. In this study, we analysed and found that when the number of transmissions are bounded by constants ≤ 20 , the likelihood of successful broadcast is small. Using simulations, we observed that the diameter decreases very fast as the transmission radius increases. Another observation is that the largest connected component emerges when the transmission radius reaches $0.3\sqrt{A}$, where A is the area containing the nodes. This may be used to determine the ideal amplification, although further simulations on larger networks could be helpful. We also found a large gap between the number of nodes required to populate the area, when all the nodes must be connected, or when only 90% of the nodes are connected.

CONTENTS

- 1 Introduction
- 2 COMMUNICATION TOPOLOGIES
- 3 GRAPH ANALYSIS
- 4 SIMULATION RESULTS
- 5 Conclusion

1 Introduction

In missions that explore the surface of planets, many light weight, low energy units such as landers, rovers, sensors will be used to collect data. The data is collected and forwarded back to Earth via an orbiter acting as a relay over long haul links, such as the links typically used in the Deep Space Network (DSN). Since the surface units are low energy units, it is essential to minimize each transmitting node's resources. They must be sufficient to establish a network while minimizing mutual interference and overall cost. For example, if the nodes of a network need to route each other's packets, then each node should ideally transmit with just enough power to guarantee the connectivity of the network.

The concept of collectively utilizing distributed sensor modules in a hierarchical manner was first introduced as *cooperative sensor networking* [1]. An extension of this idea is to prolong the life-time of finite energy sources by leveraging *cooperative modulation techniques* [2]. This relies heavily on the efficient

usage of battery power on the local communication links and requires some sharing of information, which motivates our investigation into the communication topologies for energy-efficient broadcast.

The connectivity among nodes directly influences the efficiency and reliability of information dissemination within a network. Conventionally, the topology of an ad-hoc network is defined by the transmission radius r of each node. Due to the dynamic and ad-hoc nature of such networks, using a fixed r might not render a connected network at all times. Sometimes, the network is partitioned into several connected components where each component is a connected sub-network, but there are no connections between the different sub-networks; we call this a partitioned network.

In [3], Gupta and Kumar showed that, given n nodes such that each node covers an RF circular area $\pi r_{RF}^2 = \frac{\log n + c(n)}{n}$, then the network approaches connectivity with probability one as c(n) (the connectivity measure in [3]) approaches infinity, synonymous to the number of nodes approaching infinity.

In a previous study [4], we examined the alternate extreme of [3] and established that the probability of successful broadcast is low in an area with sparsely spaced nodes. Since the nodes are sparsely spaced, the probability of obtaining a connected network is low because each node has a restricted communication radius. This motivated us to consider graphs with guaranteed connectivity; we considered the minimum spanning tree (MST), the relative neighborhood graph (RNG), and the minimum radius graph (minR). The MST is a tree connecting all the nodes where the total edge length of the tree is minimized. The RNG contains edges, where each edge connects two nodes that are at least as close to each other as they are to the rest of the points (see [5] for a precise definition). Assuming each node must use the same transmission radius, a minR graph is obtained by finding the smallest radius r which guarantees network connectivity. From our previous simulations, we found RNG to be a suitable topology for energy-efficient communication because it compares favorably to MST and minR in terms of transmission radius, edge density, node degree, fault tolerance, and hop diameter. These graph properties affect the energy usage, scheduling and reliability of the network.

In a wireless network, we want to operate the amplifiers opti-

Standard IEEE copyright which, for 2002, will be "0-7803-7231-X/01/\$10.00/\$ © 2002 IEEE".

mally at saturation. On the other hand, assuming a fixed bit rate, we cannot use a fixed power level if we want to support a communication topology defined by the RNG because RNG contains edges of variable lengths. There is clearly a trade-off between these objectives.

In this paper, we approach the problem of balancing the above mentioned objectives by studying the connectivity structures of the communication graphs, assuming a fixed communication radius for the nodes. Specifically, we are interested in the lower bound on the number of transmissions required for broadcast from a node to all the other nodes in the network.

We derived graph theoretical analyses, considering two categories of graphs. The first category consists of graphs whose connectivities are independent of the spatial distribution of the nodes. The second category contains graphs whose connectivities are influenced by the spatial distribution of nodes. Trees are considered in our analyses as a special case of connected graphs. We obtained a lower bound on the probability of trees having a hop diameter d. Upper and lower bounds on the probability of connected graphs having a hop diameter d are also derived.

For our simulations, we generated n nodes randomly placed in a fixed square area, and computed the communication graph topology assuming a transmission radius of r, where n and r are variables. The resulting data from the simulations are plotted and compared to the plots produced from our graph analyses. We observe similarities in the theoretical and experimental results.

2 COMMUNICATION TOPOLOGIES

The diameter, or hop diameter, of a graph is the maximum number of hops between any pair of nodes using a shortest path connecting the nodes. Usually, we expect a larger diameter for sparse graphs and a smaller diameter for dense graphs. In extreme cases, the diameter of a sparse graph with n nodes can be as high as n-1 while the diameter of a fully connected dense graph is one. However, as [6], [7] pointed out, many real-world graphs are sparse, but their diameters are around $\log n$. An example is the World Wide Web. The number of edges in this real-world graph is closer to n, than to $\binom{n}{2}$. Yet, it was reported that the diameter of the Web is 19 [6]. The reason that such a enormous graph has such a small diameter is because there are short-cut links each spanning a large distance, and each such link contributes only one hop to the diameter. In these graphs, the topologies of the graphs are independent of geometric distances and the spatial relationship among the nodes. That is, it may be as likely for a node to connect to its nearby neighbors as it is for the node to connect to distant nodes. In this sense, the Web is similar to random graphs [8]. In a random graph G(n, p)with n nodes, each possible edge has the same probability p of being chosen. These graphs tend to have a regular structure and a diameter proportional to $\log n$.

However, random graphs may not be an appropriate model for sensor networks. In a sensor network, each node can only communicate with other nodes within a bounded radial area. Thus, we need to incorporate geometric dependencies and spatial relationships when studying the topology of sensor networks. We are especially interested in the diameter of sensor networks because this can be used as a lower bound on the number of transmissions needed for broadcast. Intuitively, the diameter of a sensor network might be higher than $O(\log n)$ because the transmission radius could restrict the edge lengths, producing longer paths between

3 GRAPH ANALYSIS

Our graph analysis is based on several sets of assumptions.

Spatial Independence

First, we consider n randomly placed nodes in an area A, where A is a variable and its value is a finite number. The nodes are assumed to have unique positions, so two nodes cannot share a common location. The unique position of a node can be used as the node's unique label. Therefore, our analysis concerns labeled graphs. In order to broadcast from a node to all other nodes successfully, connectivity of the graph must be guaranteed. Let $Enum_{conn}(n)$ and $Enum_{all}(n)$ denote the number of connected labeled graphs and the number of all possible labeled graphs on n nodes respectively. The probability of connectivity for a labeled graph is then $Enum_{conn}(n)/Enum_{all}(n)$. Combinatorially, each edge is either included or excluded from the graph. In other words, each edge has a 50% chance of being included. Thus, $Enum_{all}(n) = 2^m$, where m is the number of all possible edges. We have n nodes, so n * (n - 1)choices of node pairs are possible. Since the node pair (x,y)and the node pair (y, x) are represented by the same nondirected edge, the number of all possible edges is $\frac{n*(n-1)}{2}$. So, $Enum_{all}(n) = 2^{\frac{1}{2}n*(n-1)}$. Indeed, this is equivalent to [9, Theorem 15.1] stating that the number of labeled graphs with n nodes is $Enum_{all}(n) = 2^{\binom{n}{2}}$. We know that a minimally connected graph is a spanning tree. Thus, the number of labeled spanning trees serves as a lower bound on the number of connected labeled graphs. According to [9], the number of labeled trees with n node was reported in [10] to be n^{n-2} by Caley over a century ago. This can be used as a lower bound on $Enum_{conn}(n)$.

Bounded Diameter Trees—Motivation to obtain a closed form solution for the MST leads to the following result.

Theorem I (Weak lower bound diameter likelihood on trees). For n nodes, the probability of generating a tree with diameter d is lower bounded by $\frac{\binom{n}{d+1}(d-1)^{n-1-d}}{n^{n-2}}$.

If $n \ge 2d - 1$, then we have a tighter lower bound of $\frac{LB_{tree}(n)}{n^{n-2}}$ where

$$LB_{tree}(n) = \binom{n}{d+1} \sum_{i=1}^{d-1} (d+i-2)^{n-d-i} . \tag{1}$$

Proof of Theorem 1: Consider a connected labeled tree with n nodes. Assume that the diameter of the graph is d where

 $2 \le d \le n-1$, and fix a node such that the diameter of the graph contains the node as an end-point; root the tree at this end-point.

Since there are n-1 nodes that are connected in some configuration, for diameter d, we have d+1 nodes representing the diameter. We consider the path representing the diameter of the graph as the core of the tree. Thus, on the core, there are d-1 possible locations for the n-1-d remaining nodes, where the n-1-d nodes are placed with direct connections to the d-1 nodes on the core.

Thus, since each of the n-1-d nodes can be placed in any of the d-1 locations, we have a $(d-1)^{n-1-d}$ scenarios for a diameter d where all cases involve nodes directly connected to the d-1 nodes on the core.

This quantity does not factor in multi-hop extensions off the first d-1 nodes. Thus, we can tighten this lower bound for branches extended off the core such that the number of possible node locations increases by one while the number of available nodes reduces by one. Specifically, we have $\sum_{i=1}^{d-1} (d-1+i-1)^{n-1-d-i+1} = \sum_{i=1}^{d-1} (d+i-2)^{n-d-i}$. However, we require $n \geq 2d-1$.

We can choose from any of the n nodes to construct the initial d diameter, resulting in $\binom{n}{d+1}$ possible scenarios.

Finally, from [10] we have the total number of trees is n^{n-2} and we are done.

As depicted in Figure 1, we see that the lower bound probability is a weak bound for a large number of nodes. However, the plot does provide some insight into the relative likelihood of the graph topology with respect to a graph diameter.

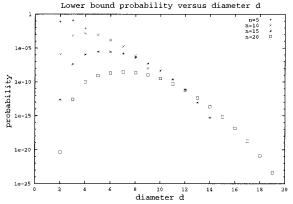


Fig. 1. Lower bound on the likelihood of a a graph diameter for n nodes.

We do not know how to compute the exact number of $Enum_{conn}$, here we suggest an upper bound.

Bounds on Connected Graphs— A necessary condition for a graph G with n nodes to be connected is that G must contain at least n-1 edges. Thus, if we subtract the number of graphs with

less than n-1 edges from $Enum_{all}$, we get an upper bound for $Enum_{conn}$. Similarly, we can extract all the cases of the n-2 nodes connecting to at least one of the initially connected nodes to obtain a lower bound on the number of connected graphs for n nodes. We obtain,

$$B_{upper}(n) = 2^{\binom{n}{2}} - \sum_{i=1}^{n-2} \binom{\binom{n}{2}}{i},$$

and

$$B_{lower}(n) = 2^{\frac{n(n-1)-2}{2}} \prod_{k=1}^{n-2} (1 - \frac{1}{2^{k+1}}).$$

Lemma 2 (Upper and Lower bound) . For n nodes there exists an upper bound, $B_{upper}(n)$ and lower bound, $B_{lower}(n)$ on the total number of connected graphs.

Proof of Lemma 2: Note that, subtracting the number of graphs with less than n-1 edges from $Enum_{all}(n)$, we get an upper bound for $Enum_{conn}(n)$. This is an upper bound because having n-1 or more edges is a necessary but not a sufficient condition for connected graphs. The number of graphs with n nodes and k edges is $\binom{\binom{n}{2}}{k}$ according to [9]. Thus, we have

$$Enum_{conn}(n) \leq 2^{\binom{n}{2}} - \sum_{i=1}^{n-2} \binom{\binom{n}{2}}{i}.$$

For n nodes, suppose we connect two nodes. There are 2 ways to connect the third node to each of the first two nodes and one way to connect to both nodes, or rather, we have $\binom{2}{1} + \binom{2}{2} = 2+1=3$. We can generalize this such that for each of the n-2 remaining nodes, the number of possible connections increases by one where the k^{th} node, for $k=1,\ldots,n-2$, has a total of $\sum_{i=1}^{k+1} \binom{k+1}{i}$ possible connections to the previously connected k nodes.

Recall the binomial theorem for x + y raised to the n^{th} power may be represented as

$$(x+y)^n = \sum_{i=0}^n \binom{n}{i} x^{n-i} y^i.$$
 (2)

Recognize that the total number of possible connected graphs $Enum_{conn}(n)$ is lower bounded by all possible connections for the other n-2 nodes, and then using (2) where x=y=1, we have

$$Enum_{conn}(n) \geq \prod_{k=1}^{n-2} \sum_{i=1}^{k+1} {k+1 \choose i}$$

$$= \prod_{k=1}^{n-2} (2^{k+1} - 1)$$

$$= 2^{\frac{n(n-1)-2}{2}} \prod_{k=1}^{n-2} (1 - \frac{1}{2^{k+1}}).$$

Corollary 3 (Weak bound on graph diameter likelihood) . For n>2d-1, the probability of a graph with diameter d is lower bounded by $\frac{LB_{tree}(n)}{B_{upper}(n)}$.

Proof of Corollary 3: This can be derived from (1) of Theorem 1 and the upper bound from Lemma 2.

Spatial Dependence

Motivated by the work of Gupta and Kumar [3], work by Jennings and Okino [4] examined the likelihood of the number of transmissions, for a relatively sparse network. The plot from [4] for the number of transmissions provided that all the nodes n reside in an area A_n such that

$$A_n \geq \frac{6\pi + (2\pi + 3\sqrt{3})(n-2)}{6}r^2,$$
 (3)

where r, the maximum communication radius, of a node is shown in Figure 2. We see that there is striking similarity between the likelihood of such events and the spatial independent plot of Figure 1.

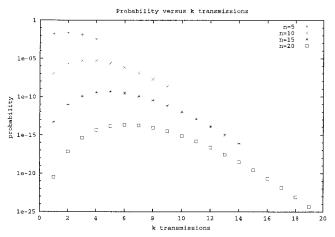


Fig. 2. The likelihood of successful broadcast given the number of transmissions and no knowledge on connectivity for various number of nodes n

4 SIMULATION RESULTS

For our simulation runs*, we generated n nodes, randomly placed in an area A, where $n=5,10,15,\cdots,100$. and A is a fixed area of 600^2 square units, and diagonal $600\sqrt{2}$ units. Two uniformly distributed random integers are generated as the coordinate of each node. The transmission radii are chosen as a fraction of the length/width of A. The factors used are: $.01,.02,.05,.1,.15,\cdots,.85$. Since the length/width of A is 600, the specific radii r used are: $6,12,30,60,90,\cdots,510$. For each n, we generate 1000 node sets. Each node set is paired with each r to obtain a graph G. Then, we compute the number of connected components in G. If G contains a single connected

component, then the graph is connected and we compute the diameter of G. If G contains more than one connected component, then we choose the largest connected component and compute its diameter.

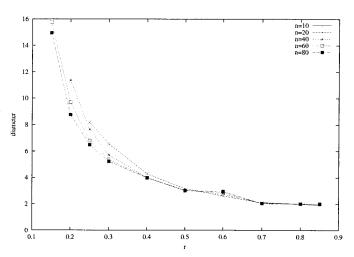


Fig. 3. The hop diameter of the graph for specific n with respect to r.

Figure 3 shows the hop diameter of a connected graph with n nodes, averaged over the 1000 graphs. Clearly, the hop diameter decreases as r increases. Note that, when n is large, we obtain a connected graph at smaller r values. An interesting question is, "for a given radius r, what is the minimum number of nodes required to achieve connectivity?" In Figure 4, we plot the smallest n values such that the number of nodes in the largest connected component is $\geq n \times v$, where v = .5, .75, .9, 1.0.

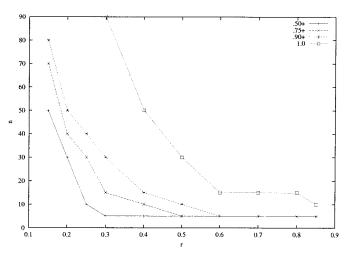


Fig. 4. Minimum number of nodes required to achieve partial connectivity.

Note that, to guarantee connectivity at r=0.3 requires ≥ 90 nodes. On the other hand, for the same r, we only need to have ≥ 30 nodes to guarantee that 90% of the nodes are connected. So, for applications that do not require all the nodes to be connected, we can use fewer nodes. In the above example, we can cut down the number of nodes to one-third.

^{*}We have implemented the algorithms in JAVA (version 1.2) on a Sun Ultra-10 workstation.

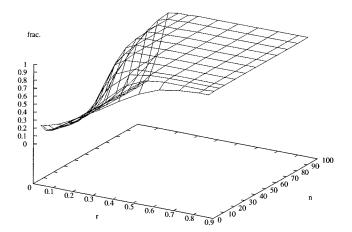


Fig. 5. The fraction of nodes in the largest connected component in graphs produced by each (n, r) pair.

From Figure 5, we observe that at r=0.3, a large connected component emerges very fast as n increases. When r>0.3, even with a small n, we still get a large connected component. So, for randomly place points using uniform distribution, it seems that if we use a radius fraction of 0.3, there is a high probability of connectivity when $n\leq 100$. We believe that a smaller factor of r can be used as n becomes larger. However, we need to run further simulations to verify this.

Using fixed r values, we observe the largest connected component with respect to hop diameter and n. Suppose the largest connected component has i nodes, we define *connected portion* as $\frac{i}{n}$. We used 14 different fraction values for r, but we choose to include only four representative plots.

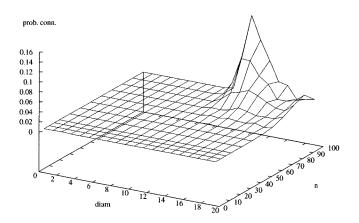


Fig. 6. Connected portion with r = .15.

Note that Figures 6-9 all have the similar shape as Figure 1

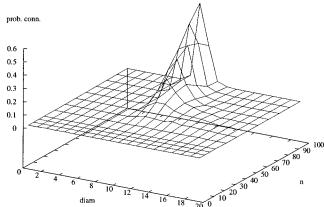


Fig. 7. Connected portion with r = .2.

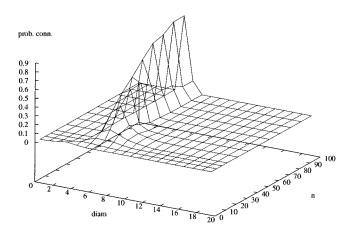


Fig. 8. Connected portion with r = .3.

where there is a single peak. Figure 6 has a close resemblance to Figures 1 and 2 because as n increases, the highest connected portion tends to have a smaller diameter value. This can be seen by regarding the curves along the x-axis for the different values of n, and see the highest point shifts to the left as n increases. It is also interesting to observe that as $r \geq 0.2$, the highest point appears to occur on the same diameter value, as shown in Figures 7-9. This seems to imply that the diameter of the largest connected component is around the same value regardless of the values of n. It also implies that r is large enough so that the diameter stays at a small constant. In Figures 6-8, we observe that the connected portion approaches one as n increases. Figure 9 shows that if we use a large $r \ge 0.8$, then the connected portion is one no matter what n is. It also show that the diameter is at one which means the graph is similar to a clique (fully connected graph).

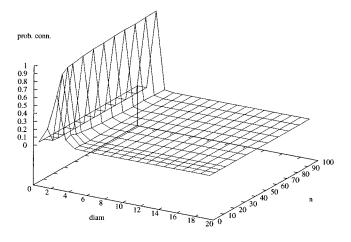


Fig. 9. Connected portion with r = .80.

It seems discouraging that the connected portion for small r is so low. However, our simulation runs generate at most 100 nodes in an area of 360,000 square units. That means, we have one node per 3600 square units, which has a diagonal of $60\sqrt{2}$ units. To determine what fraction this is with respect to a side of A, we get $\frac{60}{600}\sqrt{2}=0.1\sqrt{2}\approx 0.1414$ which is very close to the r value we used 0.15. Expanding on this, if we increase n, we conjecture that graph connectivity will occur at r<0.15. On the other hand, we expect the diameter to increase with respect of n.

5 CONCLUSION

The desire to efficiently reduce the overall energy-per-bit of a node motivated this study on the diameter of sensor networks. Diameters are important because it can be used as lower bound estimates on the number of transmissions required to broadcast information. We derived a lower bound on the likelihood of a tree with n nodes and diameter d, in terms of d and n. We also derived an upper bound and a lower bound on the likelihood of a connected graph with n nodes and diameter d. From these analyses, we estimated the likelihood of successful broadcast in a network of n nodes where the number of transmissions is fixed. We observe that when n reaches 20, the probability of successful broadcast within a fixed number of transmissions becomes very small.

From our simulations, we observed that the diameter of a network decreases very fast as the transmission radius increases. Specifically, as r is increased, the diameter is decreased by a fraction. We also noticed that at r=0.3, we can obtain a connected graph with as few as 10 nodes. If guaranteed connectivity of all the nodes is not required, we found that fewer nodes are needed to populate the area to have $\geq 90\%$ of the nodes being connected. In this initial study, the result seems to suggest that a large connected component emerges at r=0.3. Another interesting property observed is that the largest connected com-

ponent seems to have the same diameter regardless of n.

For future research, we intend to run additional simulations on larger n values to examine whether a large connected component would emerge at smaller values of r. We will also examined whether the largest connected component will still have the same diameter for larger values of n. Another interesting aspect is the relation between n and the diameter. Specifically, we wish to know that given n randomly placed nodes, could we expect the diameter of the network to be logarithmic with high probability.

As another extension to our current work, we propose a new metric representing the confidence on connectivity which we call connectedness. The concept of connectedness is in the spirit of the work of [11] on currentness for web page access rates and reliability of information. For n nodes, we say that the largest component contains $v \cdot n$ nodes for transmission radius r with probability α . Specifically, A network is said to have (n, v, α, r) connectedness if the largest component $v \leq n$ nodes are connected with probability α using transmission radius r, and n is the potential number of nodes in the network. The best case scenario of connectedness is depicted in Figure 4 where α is not explicitly characterized and left for future work.

REFERENCES

- [1] Jonathan R. Agre and Loren Clare, "An integrated architecture for cooperative sensing networks," *Computer*, pp. 106–108, May 2000.
- [2] Kevin Quirk, Meera Srinivasan, and Jonathan R. Agre, "Cooperative modulation techniques for long haul relay in sensor networks," Nov 2001, To be presented at GLOBECOMM 2001.
- [3] Piyush Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming. Edited by W.M.McEneany, G. Yin, and Q. Zhang, pp. 547-566. Birkhauser, Boston, 1998
- [4] Esther Jennings and Clayton Okino, "Topology for efficient information dissemination in ad-hoc networking," July 2001, submitted.
- [5] Kenneth J. Supowit, "The relative neighborhood graph, with an application to minimum spanning trees," *Journal of the ACM*, vol. 30, no. 3, pp. 428–448, 1983.
- [6] Brian Hayes, "Graph theory in practice: Part i," American Scientist, vol. 88, no. 1, pp. 9-13, Jan-Feb 2000.
- [7] Brian Hayes, "Graph theory in practice: Part ii," American Scientist, vol. 88, no. 2, pp. 104–109, Mar-Apr 2000.
- [8] P. Erdös and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.
- [9] Frank Harary, Graph Theory, Addison-Wesley, 1969.
- [10] A. Caley, "A theorem on trees," Quart. J. Math, vol. 23, pp. 376–378, 1889.
- [11] Brian Brewington, Observation of changing information sources, Phd thesis, Dartmouth College, June 2000.

Clayton Okino received a BS in Electrical Engineering at Oregon State University in 1989, and a MS in Electrical Engineering at Santa Clara University in 1993. From 1989 to 1994, he was a Member of the Technical Staff at Applied Signal Technology Sunnyvale, CA. In 1998, he received a Ph.D. in Electrical and Computer Engineering from the University of California, San Diego. Since 1998, he has been an assistant professor in Thayer School of Engineering at Dartmouth College. His research interests are in communication networks and wireless networks with emphasis in performance and security. His other

interest include intelligent sensors.

Esther Jennings received her BS in Computer Science at the University of California, Riverside, in 1982. In 1991, she received her MS in Computer Science from Lund University, Sweden. She received a Ph.D. in Computer Science from the Lulea University of Technology, Sweden, in 1997. From 1997-1999, she was a postdoctoral fellow at the Industrial Engineering Department at Technion, Israel Institute of Technology. From 1999-2001, she has been an assistant professor at the Computer Science Department of California State Polytechnic University, Pomona. She is currently a member of technical staff at the Jet Propulsion Laboratory. Her research interests are in distributed graph algorithms, reliable multicast protocols, scheduling algorithms in optical switches and energy-efficient algorithms for wireless networks.